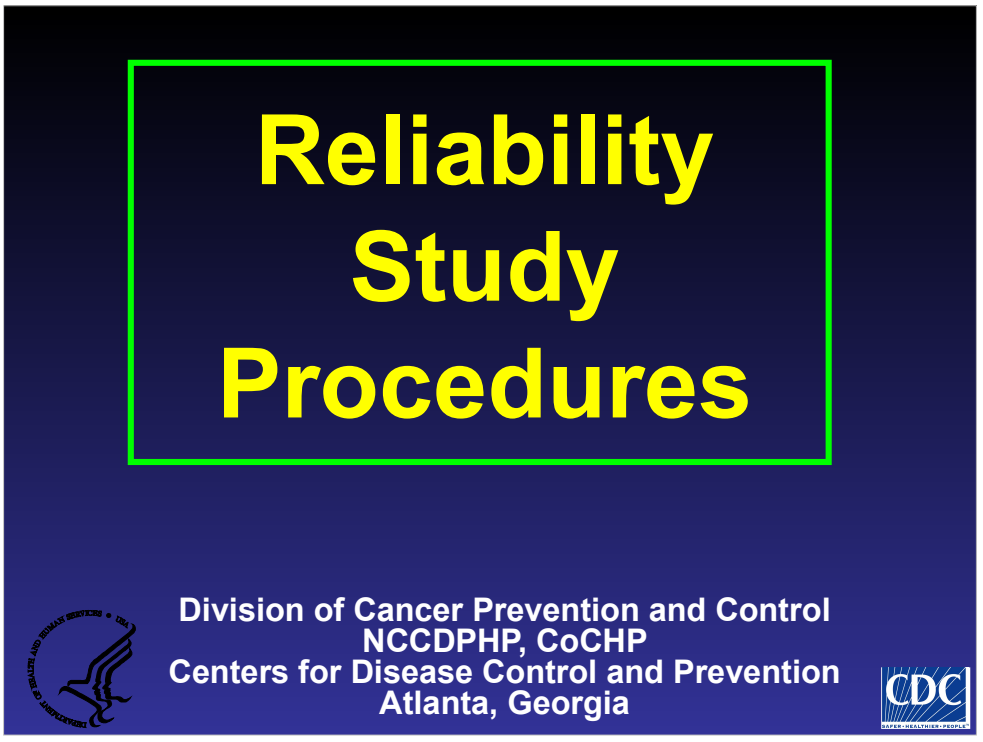
NPCR NATIONAL PROGRAM OF CANCER REGISTRIES

NPCR Education and Training Series (NETS)

Module 3: Quality Control for Central Registries

Part 4: Reliability Study Procedures

Prepared by Scientific Applications International Corporation (SAIC)
CDC Contract Number 200-2002-00576-0004
Centers for Disease Control and Prevention
National Program of Cancer Registries
Atlanta, Georgia
www.cdc.gov/cancer/npcr



NOTE TO PRESENTER: This presentation is a subset of the general quality assurance module that specifically discusses reliability issues and study procedures. Some of the slides and speaker notes are derived from the general module; other slides expand on those concepts as they apply to quality improvement through reliability studies.

Registrar Skills Study

- Detect differences in data quality per data collector
  - Inexperienced vs. experienced
  - Formally trained vs. on-the-job
  - Understanding new rules or procedures
- Methods
  - Reliability studies
  - Training

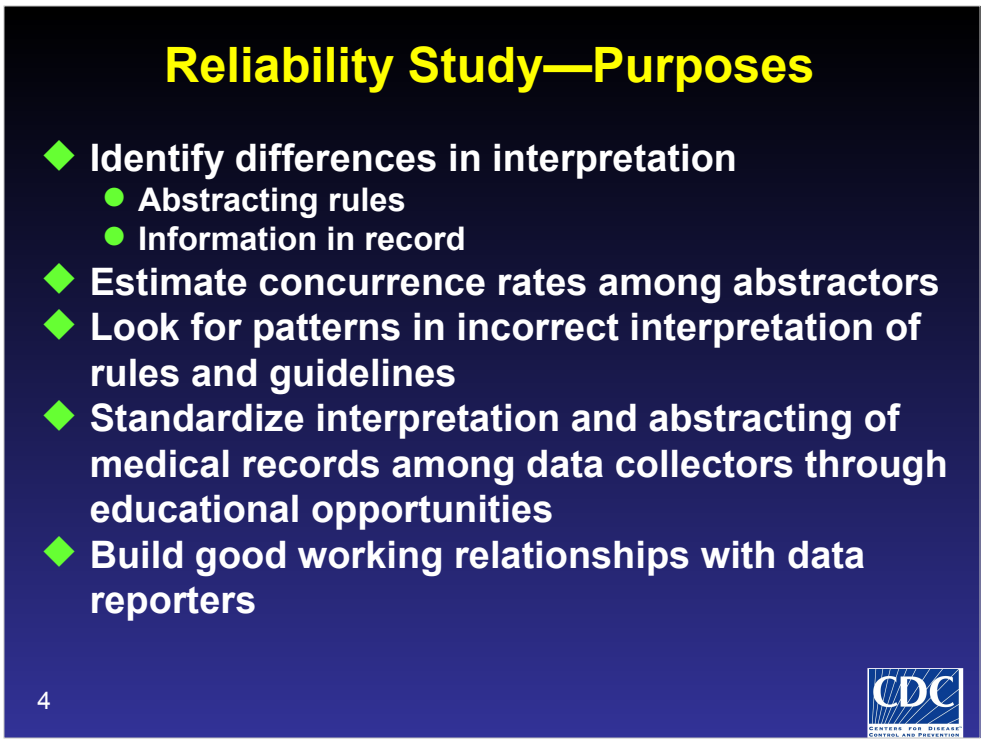
3

There is no doubt that the quality of data submitted to the central registry varies according to the skill level, training, and experience of the data collector. The skills in individual data collectors can be assessed in several ways. The central registry may design a study to compare abstracting results between inexperienced (non-CTRs or abstractors with less than two years of experience) with experienced abstractors, or formally trained abstractors versus those who learned on the job. Although this distinction between abstractors may not be a conscious one on the part of the central registry, it is a factor in validating the quality of any registry database.

Registrar skills audits are extremely useful to assess understanding of newly implemented rules or procedures. For example, an assessment study was conducted nationally of Collaborative Staging about 22 months after the system was implemented. The findings were used to improve the documentation of rules in the CS manual and target education for data collectors nationwide.

In addition to visual editing, a type of acceptance sampling of individual abstracts submitted by the data collector, more formalized audit methods can assess quality differences among data collectors. The usual way to assess individual registrars' skills is by conducting a reliability study or audit. "Reliability" is the consistency in which more than one abstractor arrives at the same code, given the same information in the medical record.

NOTE: Throughout this presentation we may use the terms "audit" and "study" interchangeably. An audit is a more structured and formal process; a study may involve just as much work but is not as formal.



The purposes of a reliability study are similar to those of a reabstracting audit and include—

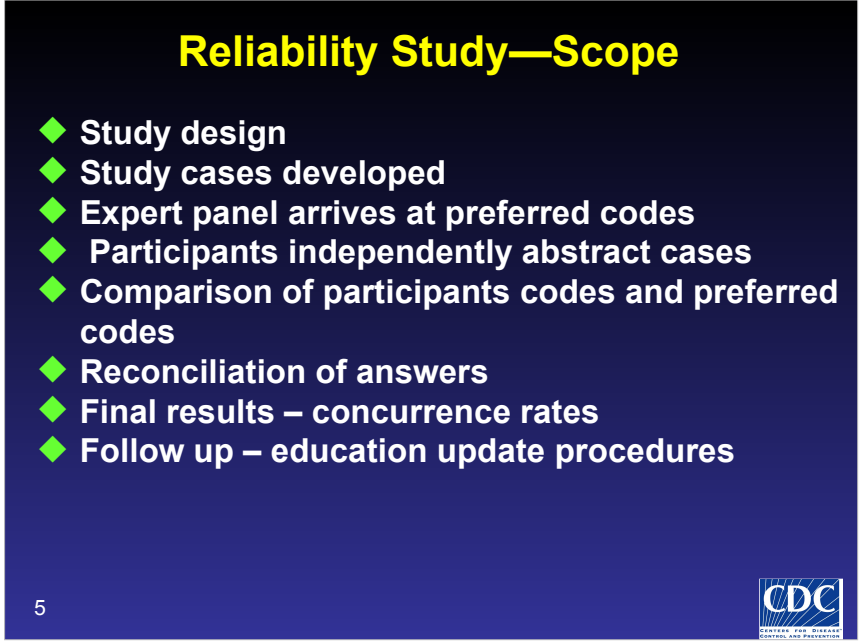
•Identifying differences in the interpretation of abstracting and coding rules, or interpretation of information available in patient records. For example, a difference in coding between two abstractors is an indication that at least one of the abstractors interprets the facts about the case differently. This may be due to lack of understanding about the facts or unclear instructions about how to code the facts.

•Estimating concurrence or agreement rates among abstractors. Have the abstractors arrived at the same code based on their understanding of the case information and rules for coding? If not, what could explain the differences?

•Looking for patterns in incorrect data that would provide opportunities for further education and training, and ultimately, overall improvement of the data.

•Standardizing the interpretation and abstracting of the medical record among data collectors through educational opportunities based on the results of the audit.

•Building good working relationships with data reporters. Reliability studies are immensely educational, since they assess understanding of rules and guidelines for abstracting and provide opportunities for abstractor training and professional development.



Reliability studies are designed to test participants' understanding and compliance with established coding rules and practices. This is the only type of study that can evaluate the overall performance of coders and abstractors. Other terms that apply to this process are concurrence, compliance, consistency, and reproducibility. The point is to assess how consistent individual abstractors are in determining a code when faced with the same set of information. Reliability studies are relatively easy to develop and can be designed to target any number of areas in the abstracting process.

This slide details the major activities involved in a reliability study.

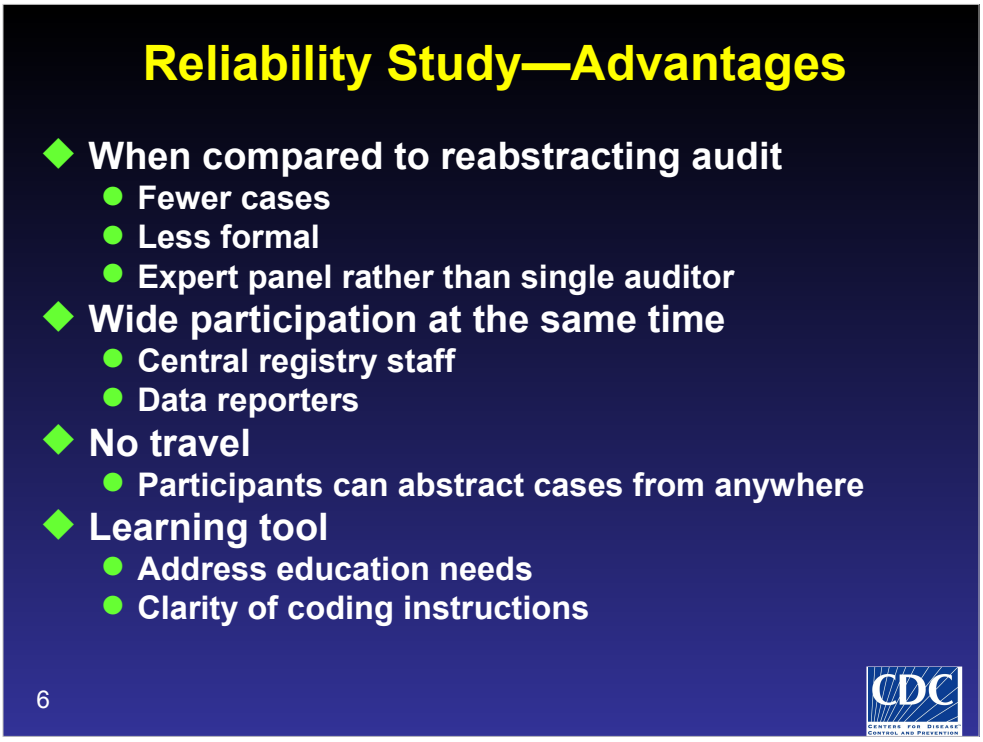
How about focused reliability on certain fields (i.e., CS staging)? Or when new fields are introduced or new staging?

Compared to a reabstracting audit, a reliability study is somewhat less formal and involves fewer cases, since all participants abstract the same test cases. Participation in a reliability study may be voluntary or required, as determined by the study team. The study must be designed carefully to ensure the validity of conclusions drawn from the results, particularly if participation is voluntary. The target subject matter and design of the study determine which and how many data fields (such as staging or treatment fields) will be abstracted and analyzed.

The principal advantage of a reliability study compared to formal audits is the cost savings on travel, no on-site abstracting is necessary. The amount of time involved in developing the study and the amount of time needed to analyze the results are at least as much for casefinding and reabstracting audits.

The study team develops sample cases (mock charts) as part of the procedure for a reliability study. These sample cases use actual or modified source documents that are based on the subject matter targeted for the study. An expert panel develops preferred answers for each of the sample cases. The expert panel must be intimately familiar with coding and abstracting rules as well as the cancer disease process and what to look for in the medical record.

Study participants then abstract the mock charts and submit their codes. The responses are aggregated and distributions of codes relative to the preferred answers are prepared for review. During the reconciliation process the participants can question the preferred answers and comment on coding and abstracting issues in the case itself. The reconciliation process will be discussed in detail later in this presentation. Finally, when all answers are finalized, the individual results are compared once more with the final answers and concurrence rates can be calculated.



Compared to a reabstracting audit, a reliability study is somewhat less formal, involves fewer cases, and all participants abstract the same test cases.

Many people can participate at the same time, both central registry staff and data reporters.

One of the principal advantages of a reliability study compared to formal audits is the cost savings on travel—no on-site abstracting is necessary.

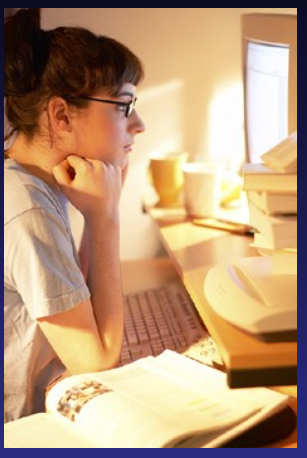
This is the only type of study that can evaluate the overall performance of coders and abstractors.

The cases in the study can indicate more widespread patterns in the understanding and application of coding rules and guidelines.

Reliability studies are a successful learning tool and can be used to address education and training needs as well as clarify coding instructions.

Reliability Study—Drawbacks

- Time commitment on part of participants
- No conclusions can be drawn about data in central registry database



7

There are a few drawbacks to reliability studies. The amount of time involved in developing the study can take longer than a traditional audit because the time involved needed to obtain or create the test cases, develop preferred answers, and conduct the reconciliation process.

Although a reliability study is cost-effective for the central registry to conduct, it requires a substantial time commitment on the part of the participants, and many more people outside the central registry are involved in the study. Carefully abstracting each mock chart in a reliability study takes as much, if not more, time as abstracting “normal” charts in the facility registry. Depending on the difficulty of the cases it could take as much as an hour or more. To successfully participate in a reliability study, abstractors must complete all of the cases. Incomplete results will affect the overall analysis and conclusions drawn.

Although registrar skills audits, such as reliability studies, take time away from the registrar’s normal work load, they are an immensely successful learning tool. The Collaborative Staging Reliability Study conducted in November 2005 had wide-ranging effects. First, it showed that registrars are willing to take the time to find out whether they are doing things the same way as other registrars. Second, it showed that there were some unclear instructions in Collaborative Staging documentation that have since been addressed in updates to the manual. Third, the CS reliability study demonstrated several areas in the abstracting process where further education was needed, for example in understanding the anatomy of the primary organ and the relationships of adjacent organs and structures. These educational issues have been addressed in a series of recorded presentations available for review on the Commission on Cancer/AJCC online education center.

Although a reliability study is useful in assessing overall quality of abstracting, it is only a measure of accuracy and understanding and/or interpretation for the cases in the reliability study. The study cases are not necessarily representative of typical cases in the central registry database, especially if the study cases have been developed to address particular data quality issues. Consequently no conclusions can be drawn about the data in the central registry database. On the other hand, the cases in the study can indicate more widespread patterns in the understanding and application of coding rules and guidelines.

Developing a Reliability Study

- Study team
- Topic selection
- Timing
- Writing the protocol
- Case selection
- Developing preferred answers
- Conducting the study
- Analyzing the data
- Feedback

8

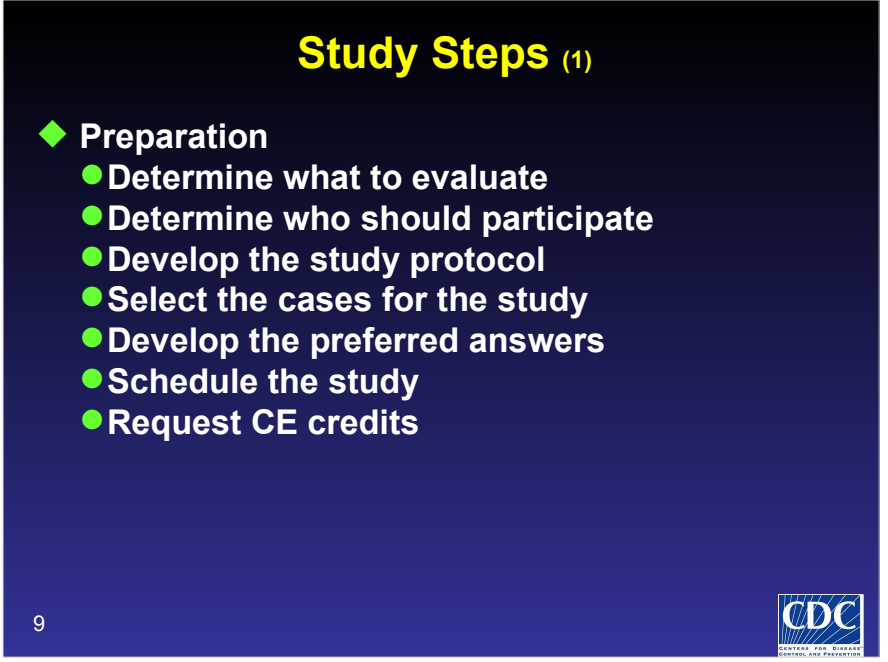
Numerous factors must be considered when any type of audit or study is developed. This is not a one-person job; it takes a team of central registry staff to do a good job designing and conducting the study. There must be documentation of the process from beginning to end, including the reference documents that will be the foundation for coding rules and guidelines used in the reliability study.

There are a variety of ways to select the topic of the reliability study, and the timing must be carefully planned to avoid overburdening both central registry and participants.

The audit process begins with protocol development and concludes when the findings have been provided to the participants. All of the aspects of conducting the study have to be carefully planned. In fact, a reliability study may take longer to develop than other types of audits because of the extra time needed to create the cases, develop preferred answers, and conduct the reconciliation process.

The project's not finished until the data have been analyzed, summarized, and plans made for follow-through on the findings.

Let us take a look at each of these steps in more detail.



The first step is to determine what to evaluate. Many aspects of the study must be considered. All of the study design and development must be completed before logistics are considered and the study actually begins.

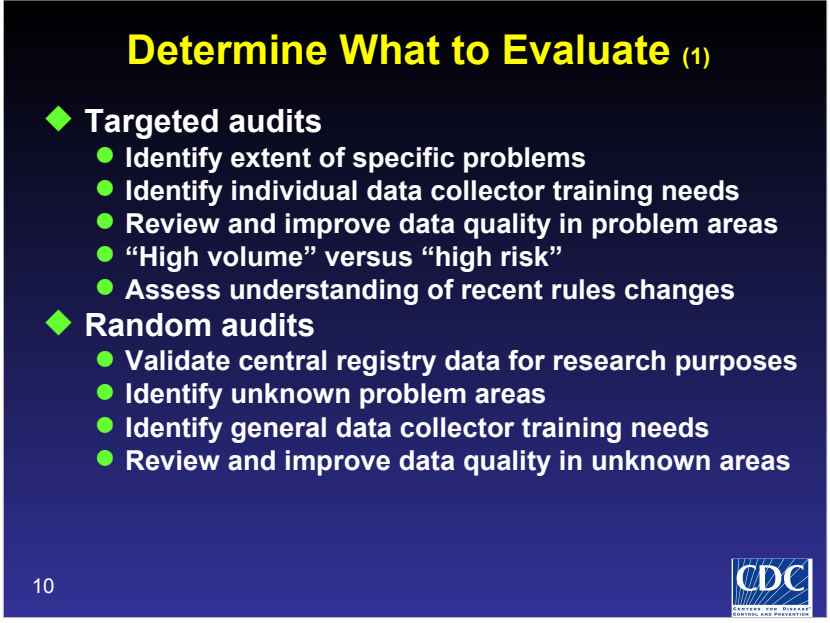
The next step is to determine who should participate and if participation will be voluntary or required. This step depends on the subject matter. For example, if the reliability study is intended to assess understanding of newly implemented rules, then anyone who abstracts a case should participate, including central registry staff who oversee data quality.

The best way to assure consistent application of the design is to develop a protocol for everyone to follow.

Once the criteria for the study have been established and described in the protocol, the cases should be selected. As the cases are being prepared, an expert panel should be selected and asked to develop preferred answers for the finalized cases.

Determining the timing of the study is the next step. Participants must be notified and may need to obtain permission to participate.

Let's look at each of these in a little more detail.



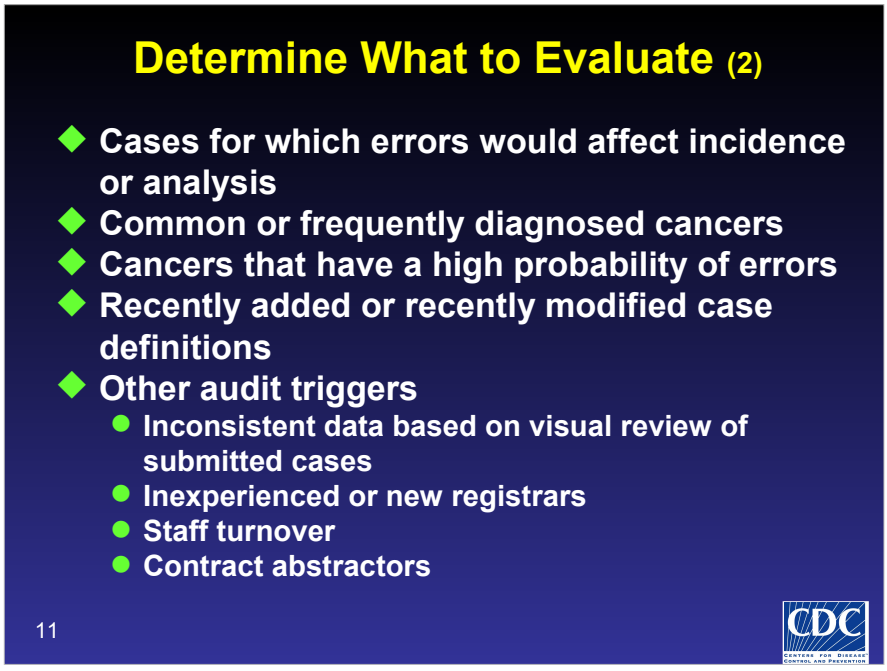
What will be the topic of the study? This is a major decision that affects many aspects of study design. In fact, the topic may be the first aspect selected. The two big categories of audits are targeted and random. The purposes of these types of audits are different. Targeted audits have been triggered by something—usually an actual or perceived problem with the central registry data. The targeted audit will help determine the extent of the problem. Targeted audits can also identify training needs for individual data collectors. The result of the audit will be increased knowledge on the part of the participants specific to the problem area under study. Training or other feedback from the audit may have wider beneficial effects because of the attention paid to the abstractors participating in the study.

As with other types of formal audits, an important concept in quality assurance methodology is to target areas of high-volume or high-risk for periodic audits. High-volume areas include those with many cases, such as the major cancer sites. Any issues identified and corrected in a high volume area will improve a large sector of the central registry database. For example, identifying problem patterns in and better training about the relationships of the CS Lymph Nodes and Site-Specific Factors fields for breast will result in better quality data for thousands of breast cases. High-risk areas are those prone to error but which do not necessarily involve large numbers of cases. For example, head and neck cases are difficult to abstract, because often the primary site is not clear and there may be multicentric tumors that could represent single or multiple primaries. A reliability study covering various aspects of head and neck cancer, such as determining the primary site, multiple primary rules, staging of the disease (if applicable), and documentation of treatment would be a useful way to assess the participants' understanding of the coding and abstracting rules for these cancers. Some issues are both high-volume and high-risk, such as accurate staging of lung cancer or prostate cancer.

Reliability studies are conducive to assessing abstractor understanding of recent rules changes. The 2005 Collaborative Staging Reliability Study is an example of this, because it assessed abstractors' grasp of the new coding system implemented the year before. As previously noted, the CS reliability study discovered many areas where registrar education was needed and additional areas where coding manual instructions could be clarified and improved. The study reconciliation process forced participants to read the general instructions in the front of the manual, some perhaps for the first time.

On the other hand, random audits aren't exactly a fishing expedition. Consider a study of random data fields as a spot check or sampling of the data. It is possible that a reliability study will identify problem areas that were not previously suspected and that have not been identified through formal data quality monitoring. Another use of random audits is to obtain a clearer perspective of training needs for data collectors in general, not just those with identified problems. A random audit isn't truly random; it still must be carefully planned and executed to be meaningful and valid.

Experience has shown that because of limited resources, most reliability studies performed by central registries are targeted in one way or another. This is where it might be good to introduce ways of testing new staging, new data items, etc., conducted by standard setters to see how the new data and/or staging is understood.



The goals and expected outcomes of a reliability study should be just as clearly defined as those for more formal audits. If problems in data quality are identified, solutions must be developed and implemented to ensure that errors are corrected and future data are abstracted correctly.

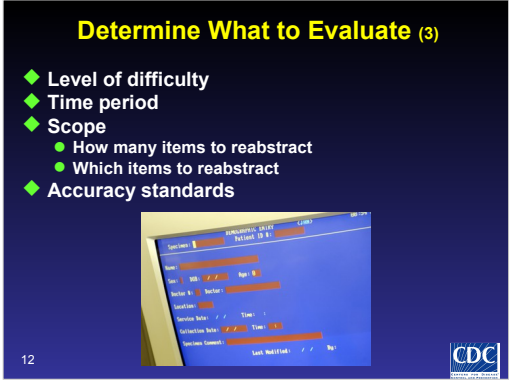
Basic principles of reliability audits include—

•Assess those cases for which errors would affect incidence or analysis. The design of the study should be based on knowledge and understanding of the central registry data and what types of data discrepancies have the most potential to affect incidence counts.

•Another option is to evaluate those cancers that have a high probability of errors. These "high-risk" targets may not be the most common sites, but could be of research interest and it is important to have confidence that the data is accurate in order to avoid incorrect conclusions by the researchers.

•One important use of the reliability study concept is to identify issues while they are still fairly easy to correct. Any type of rule change or guideline that has recently been implemented is an ideal target for a reliability study.

•Alternatively, the central registry may decide to evaluate a topic for other reasons. For example central registry visual review may have identified a data quality problem from a single facility with multiple registrars. The trigger could be an increase in the proportion of unknown or blank values in key data fields. Other targets might be inexperienced registrars, hospitals with recent turnover, or facilities using contract abstractors who may not be completely familiar with unique reporting requirements for the state.



As part of the study design, the audit team must deal with one of the unique features of a reliability study—deciding the level of difficulty of the cases. This decision is a function of the topic selected and the desired outcomes. If the central registry wants to assess general understanding of recent rules changes, then fairly general or straightforward cases should be selected for the study. These could even be random cases that have been reviewed and already incorporated into the central registry database. On the other hand, if the registry wants to assess the scope of specific issues, such as the interpretation of one or more specific coding rules, then moderate to challenging cases that present variations on those coding rules should be selected. Another approach might be to select a series of progressively more difficult cases to evaluate at what point the participants begin to have increased discrepancy rates. As a side issue, a reliability study that is too easy will be boring for experienced abstractors, but one that is too difficult will be discouraging for newer registrars. The audit team must strike a balance for all the participants, because it may be surprising where the discrepancies occur.

The time period of the study is not the time that the study is open to participants but rather the period of the rules being evaluated. The study protocol must state what rules are to be in effect for the cases being abstracted. Ideally, the most recent set of rules should be used as the basis for the study. For example, if the reliability study is being designed to assess the multiple primary rules for urinary tract cancers, the cases selected must have dates appropriate to the rules in effect. This includes diagnosis dates in 2007 or later.

As a function of the study topic, the audit team must determine what data fields to reabstract. The choices can be a very limited data set or all data items, or somewhere in between. In general, the more items that are abstracted, especially those that may not be relevant to the purpose of the study, the longer it will take the participants to finish the case. It is better to agree on a carefully defined set of data items during the development of the audit protocol. For example, if the audit is being done to assess how well abstractors have understood and applied recent changes to histology coding rules, only those data fields affected by the rule changes should be reabstracted, including histology code, primary site, date of diagnosis, and sequence number. On the other hand, if the study is triggered by the desire to check the accuracy of a new abstractor's work, the central registry may want to review all data items. The choice of data items for the study may be determined by how the data will be used.

A reliability study need not have a single focus. If the data about a cancer site are subject to interpretation in many areas, the cases in the reliability study can be designed to evaluate many factors.

The central registry study design team may opt to include all items in a data category (sometimes called a data cluster), such as demographics, tumor description, staging, and/or treatment.

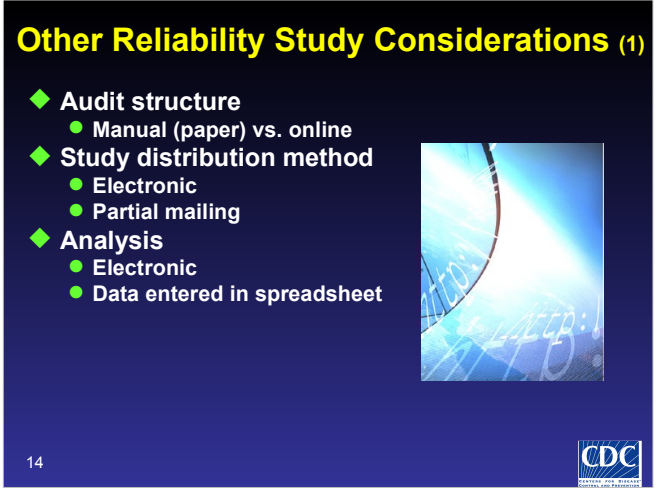
Demographics: Age, sex, race, ethnicity, and county of residence are all important factors in incidence reporting. Other demographic information such as date of birth, address at diagnosis, and social security number are important for case- and tumor-matching. Demographic information is usually very objective. The information is either there or not there in the medical record. When the purpose of the study is to identify variances in the interpretation of data, inclusion of demographic items may not be the best choice.

Tumor description: Primary site, histology, date of diagnosis, sequence of tumor, class of case, and tumor behavior can also affect incidence reporting, but they are equally important for researchers. Tumor information is usually fairly objective, but more subject to interpretation by the abstractor than demographic information. The tumor description items underwent some rules changes at the beginning of 2007 and even though they are fairly straightforward, they may be good items for a reliability study.

Staging information: The collaborative staging fields, with or without site-specific factors and the eval fields, are the way the abstractor captures the facts about the cancer and are the basis for mapping into summary stage for the epidemiologists and AJCC/TNM stage for the clinical researchers. Staging information is subject to interpretation both by the clinician and abstractor and must be evaluated carefully. Over time, the staging information fields have been shown to have the widest variation in interpretation of all the fields on a cancer abstract. Consequently they are always a good target for evaluation.

Treatment information: Date of treatment, type of surgery, type of radiation and systemic therapy are the primary data fields, but a reliability study may evaluate whether a particular type of therapy should be coded as “cancer-directed” treatment for a certain primary site. Treatment information may be subjective, especially determining first course versus subsequent treatment or whether a treatment regimen is complete, but it is less open to interpretation than staging information.

Part of the audit design is predetermining accuracy standards. At present, there are no published national standards for data accuracy. Some central registries have established accuracy thresholds over time, for example, the California Cancer Registry’s 97% accuracy expectations for 40 data items that are visually edited. The threshold for data accuracy can vary among data fields based on the relative importance of the field to incidence reporting or a specific research study. For example, sex and race are critical elements of incidence reporting; the threshold for these might be 100% accuracy. Tumor grade (when reported) is less critical to most research and might have an 80% or 85% expected accuracy rate. The threshold for data accuracy can even vary from audit to audit.



Reliability studies are unique in several other aspects as well. Casefinding and reabstracting studies deal with multiple medical records and other documents in facilities. Reliability studies work with a limited number of documents that can literally be created by the study designers. With this increased flexibility compared to other types of audits, one of the decisions that must be made as part of the study design is its structure, whether it will be conducted electronically or manually (on paper). There are advantages and drawbacks to both structures.

A manual or paper study involves minimal costs on the part of the central registry (other than development and analysis time). The study materials could be distributed either electronically or on paper. However, there are no costs for Web site development or technical support needed by participants. The principal drawback is that study answers submitted on paper must be keyed into a spreadsheet or other analysis software, thereby adding another layer of error potential in the data entry process. Of course the data could be analyzed manually, but even a spreadsheet program can offer advantages over hand calculations.

An electronic study design offers several advantages but with increased costs and some additional technical issues. If the study is designed to be conducted via the Internet, the participant can enter answers directly into the database, which then can be analyzed electronically. The cases can be posted on the Internet for downloading and review prior to beginning the data entry process, or they can be programmed to display during the study process. Online studies present some technical challenges for participants who are not familiar with the internet or whose facility IT staff has placed restrictions on internet access. For those unfamiliar with accessing the Internet, technical support from the central registry may be needed to get the participant registered and a study password assigned. Unfortunately, an online study may have to be designed for the least technical person, thereby limiting its capabilities. Costs incurred in an Internet-based study are those of programming a totally electronic study, developing an interactive Web site, and manipulating the online data for analysis.

Distribution of the study materials and the method of analysis have been mentioned already. Distribution could be entirely electronic, especially if the study is online. The documents can be created electronically and e-mailed to the participants or posted on a Web site to be downloaded by the participants. At the completion of the study, the answer sheet can be returned by e-mail or through the postal service. Increased expense is incurred when the documents are copied and mailed out by the central registry. The participant is still obligated to mail the answer sheet back to the central registry. Another possibility is to conduct the reliability study during a workshop, but full participation may be limited by the ability of the abstractors to travel to the site of the workshop. Online studies have the advantage that they are open to the participants 24 hours a day, seven days a week through the period of the study, with access from home or office. Furthermore, everyone has the same completion deadline because the Webmaster can shut down the Web site at a specific time.

Analysis of results obviously is much easier when the data are in electronic format. Study data that can be exported to statistical analysis software is ideal, and even popular spreadsheet software can do that. Avoiding a separate data entry step is a good idea and another reason that online studies are advantageous, assuming the central registry has the resources to create an online study.

Obtaining cases for the study is an important issue. As mentioned, cases could be identified from the central registry database. However, the source documents for those cases must be obtained from the submitting hospital, and there may be some release of information issues. The central registry may have access to certain types of cases that are abstracted by central registry employees, but these are not necessarily the moderate to challenging cases that make for a good reliability study. One method used by the SEER program during the CS reliability study was to announce a call for cases, requesting that hospitals submit de-identified cases that met certain criteria. SEER received hundreds of cases and selected the best ones for the CS reliability study.

One consideration during the process of announcing a call for cases is whether to request the full case, including nursing notes, medication sheets, and all other documents, or whether to request only those reports that are pertinent to the case. This latter decision places the responsibility on the person preparing the record for submission to select the reports that contain the facts about the case, but reduces the copying and de-identification burdens on the part of the person submitting the case.

Despite a wide selection of cases to choose from, the study designers may not end up with cases that meet the needs of the study. Should this be the situation, a decision will have to be made whether to alter the case to fit the desired study targets. If the reliability study is intended to be a random audit, altering the case is probably not a good idea. On the other hand, it may be a necessity if the study targets specific issues. If the decision is made to alter the facts of the case, it would be advisable to send the altered case to an impartial reviewer to make sure that one alteration did not have a domino effect on other aspects of the case. Further, the reviewer should make sure that an alteration on one report, such as a change in the segment of colon involved on an imaging study, was carried through to other reports containing that same information, such as the operative and pathology reports.

Some targeted reliability studies may not require full, detailed cases. For example, if dates of diagnosis or treatment are at issue, short scenarios (one page or less) might be created. This will save participant abstracting time and will definitely focus the study on the target topic. On the other hand, it may not be possible to truly assess the participants' ability to interpret the case information if the pertinent facts are extracted from the medical record and presented in a brief scenario. A decision to use case scenarios versus full cases will have to be a function of the target topic.

Basic Audit Principles

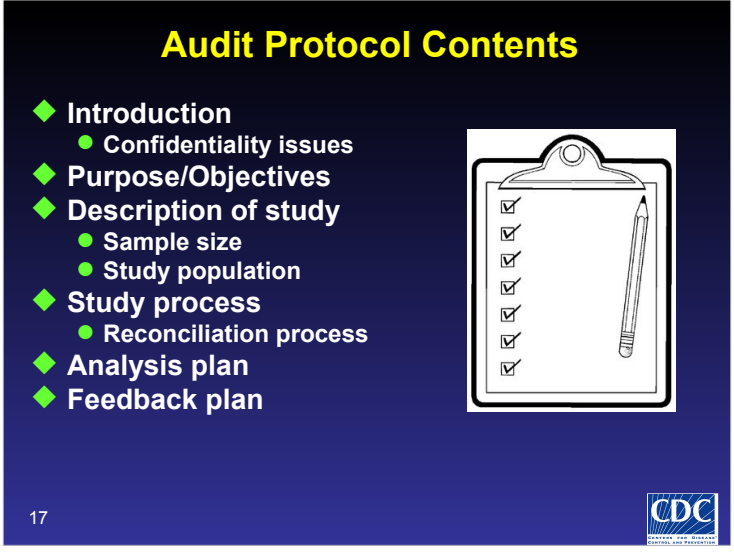
- Target the audit to get the greatest effect of available resources
- Financial considerations
  - Salaries: auditor, statisticians, programmers, support staff, and study participants
  - Communications costs
  - Audit materials (mock charts, abstracting software)

16

A basic principle of developing and conducting an audit is to target the audit to get the greatest effect of available resources. In the case of a reliability study, the available resources are primarily the valuable time of the central registry audit team and the participants. The study must be designed to have maximum effect for the commitment of time involved.

Financial considerations include salaries of the audit staff and development and use of study materials. The audit team consists of the study coordinator, statisticians, programmers—including Internet programmers if needed, and support staff. The expert panel is a tangential part of the audit team. Instead of a single or limited number of central registry auditors, there will be dozens if not hundreds of participants, each contributing hours of work time to the study. The study coordinator may call in consultants or subject matter experts as part of the study planning, development, and analysis processes. A reliability study may also involve communications costs, such as Web sites and telephone conference calls or Web conferencing during the reconciliation process.

Specialized software will be needed for the audit, so programming costs for developing the software must be part of the budget. Development of the software should begin months in advance of the audit itself. It may be possible to modify data abstraction software, but the desired reporting features of the software must be incorporated into the modifications. In a reliability study, only a limited data set is abstracted, so the software must have the capability to include or exclude data fields, especially if the software is to be used year after year as audit requirements change. Additional, analyzable data fields should be included in the audit software to record discrepancy resolution comments and revised codes. The software should also be able to produce reports and possibly print both the abstracted codes and preferred answers for use in the reconciliation process. Ideally, the discrepant fields should be flagged or highlighted by the software to make the resolution process easier. Export capabilities to statistical analysis software should also be built into the software as it is being developed or modified. The SEER Program has developed internet-based reliability software, and several states have analysis software for reliability studies.



As we have said repeatedly, any audit must be designed carefully. The reliability study protocol is perhaps the most important document in the audit process, because it provides everyone with a standard set of instructions. The protocol should have clearly stated objectives, a description of the sampling plan, and an outline of what is to be looked for in the analysis. All of this should be included in a written protocol.

The protocol assures consistency among participants both at the central office and in their facilities. A thorough reliability study protocol will contain the following sections—

- **Introduction:** Provides authority to conduct audit and rationale for conducting audit, if this is needed. Any HIPAA issues or other considerations of patient confidentiality should be addressed while the audit is being planned. In general, however, the sample cases for a reliability study are scrubbed of any personal identifiers, so HIPAA should not be an issue, and this should be stated in the protocol.

- **Purpose/Objectives:** State the reason for the study, for example, to assess general knowledge of new coding or staging rules. Indicate the expected outcome of the study (error rates, better understanding, database cleanup). See analysis plan below.

- **Description of Study:** Identifies the type of study. This will give the participants an understanding of the study process. If possible, explain how a reliability study is different from a formal audit. Describe the sample cases in general terms, and provide an estimate of how long the study will take. Describe the types of participants, and be sure to explain whether participation is required or voluntary. Include instructions on registering or enrolling for the study, such as filling out an questionnaire or obtaining a password.

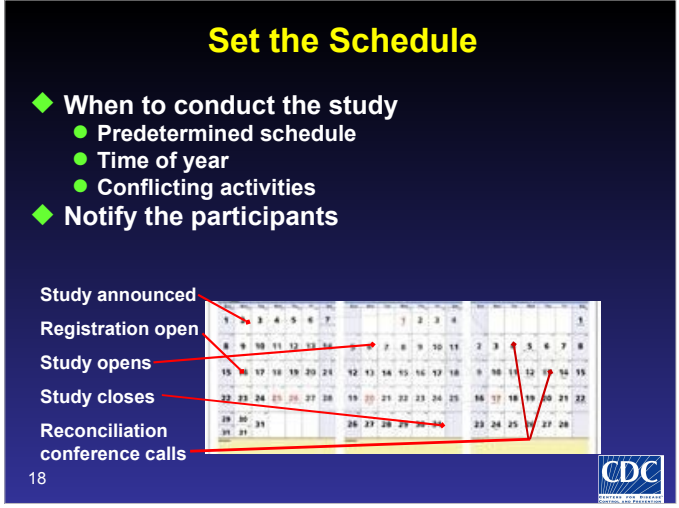
- **Study Process:** Describes when and how the audit will be conducted (via Internet, mail-in documents, in-person workshop), and the steps involved, including what reference materials may be used and how any discrepancies will be reconciled after the audit is completed.

- **Analysis Plan:** Describes what type of calculations will be part of the final report. If possible, provide templates of the analysis tables designed by the statistician member of the audit team. This too will give the participants an idea of what the study is looking for.

- **Feedback Plan:** Provides in the protocol a list or description of the final documents from the audit, such as a certificate of participation or continuing education credits. This might include an indication of whether the participants will receive a final report or just an individual overall score, a summary report or a detailed list of findings, and any plans for education or training based on the findings.

Although the protocol describes the process in detail, it is also a good idea to provide a checklist for the facility to follow, complete with due dates and references to specific sections of the protocol.

Allow as many people as possible to review the study protocol before it is distributed to the participants. All members of the central registry audit team should read the protocol to check that, if followed, the protocol will give the desired result. In addition, it is a good idea to have one or more potential participants (who may or may not actually participate) review the document, because there may be issues with access to the internet or scheduling problems that the central registry may not be aware of.



As previously noted, audits and studies are time-consuming and expensive to conduct. They disrupt the normal activities of both central registry quality control staff and the participants. Careful planning and continual communication with the participants are very important.

Decide when to audit. Is this study part of a regular audit schedule defined in registry operations or standards? If so, the study may be conducted on a predetermined schedule, annually or more often. Reliability studies provide some variation from routine, on-site casefinding or reabstracting audit conducted by the central registry. There is usually less preparation on the part of the participant, compared to requesting documents from other departments for a casefinding or reabstracting audit. The audit schedule will most likely depend on resources such as funding for study development, data analysis, and availability of staff.

An early factor in scheduling a study that requires a substantial commitment of time on the part of the participants must be the time of year that the on-site audit takes place. All things considered, there are only limited periods of time that an audit can take place. Usually the two to three months prior to a data submission are not good because everyone is involved in final database cleanup prior to the submission. Since no travel is involved for a reliability study, winter months might be an option, taking into account major holidays and end of year “use or lose” vacation policies. April, May, and June are not always good choices because of the large national meetings, including NCRA, NAACCR, and spring state and local professional meetings. Some courtesies must be given to the facilities themselves, such as avoiding Joint Commission and Commission on Cancer surveys and the period right before those surveys. Facility staff vacations and maternity leave and other factors like computer conversions or installations must be considered as well. In contrast to the relatively limited burst of activity by central registry staff to physically conduct a casefinding or reabstracting study in 10 or 15 facilities in a short amount of time, the central registry needs to plan to have the reliability study open for several weeks in order to permit participants to complete the study with minimal disruption of their normal work.

Once the period of the study has been established, the participants must be notified. The communication should indicate that the participants have been selected for a reliability study during a particular time. A copy of the study protocol and a checklist of preparation activities (if any) should be provided. If the study is held after a state meeting, the central registry should request time on the agenda to promote participation in the study. If the central registry or state association has an e-mail bulletin, the study should be announced in that as well.

The communication should also reiterate any ground rules for the study, such as a reminder that the study is conducted on the honor system, that participants may not discuss the study cases among themselves until all participants have completed the study, or that there should be no group review or preparation sessions prior to the study that might bias the results.

Study Steps (2)

- Keep promoting the study via e-mail or list-serv announcements
- Follow up with participants who have not registered or who have questions about practice cases (if provided)
- Conduct the study

19

The second part of the study process is to maintain contact with the participants regarding the specifics of the study as the opening date approaches. This is particularly important if the participant has to do any preparatory activities, such as completing a questionnaire, testing access to a Web site, obtaining a password, or working with a practice case. As previously noted, the participant should have a copy of the protocol and a checklist of reference materials to have ready.

If the participant has any problem with study access or doesn't understand the study process, these issues should be resolved before the study opens.

After the preparations have been made, it will be time to open the study to the participants. There will probably be a rush of questions on the first day, then activity will taper for awhile. Experience and human nature have shown that study participants tend to wait until the last few days of the study period to complete it. Well-designed reliability software can monitor the progress of study participants as they complete a series of cases. The study coordinator or another administrator at the central registry should monitor participant progress and "nudge" registered participants who have not begun the study by the halfway point.

Study Steps (3)

- Run the preliminary analysis
- Resolve discrepancies (reconciliation)
- Compile final results
- Prepare report with recommendations and/or deficiencies
- Distribute report
- Identify educational needs

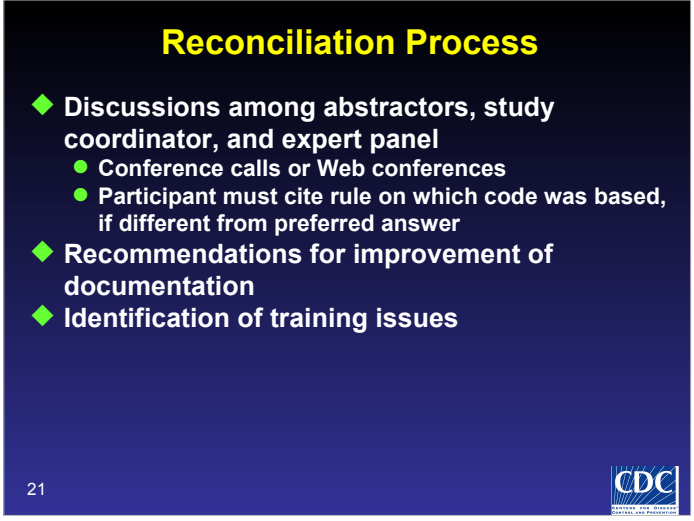
20

The final part of the study begins with completing the data analysis after the study period closes. This is a multi-step process because the analysis must be run twice, first using the expert panel's preferred answers, and after reconciliation, using the finalized answers. These final steps in the study involve resolving any data discrepancies through the reconciliation process, tabulating the final results of the study, completing the final report, determining any recommendations, and identifying any educational needs based on the audit.

For the resolution phase of the study, the analysis software should provide not only a list of data fields for each case showing the participant's codes and the preferred codes, but also a distribution of the aggregated participant codes. Through one or more conference calls, the study coordinator, expert panel, and study participants can work through the data field discrepancies and determine what is the best code for the case. More on this in a moment.

Once the best codes for each case have been determined, the participant data is run again against the finalized answers and error rates—or conversely, concurrence rates—can be calculated and disseminated to the participants. Results can be provided in a technical document or a basic overall score can be reported.

Plans for any educational efforts resulting from the study should also be determined and publicized.

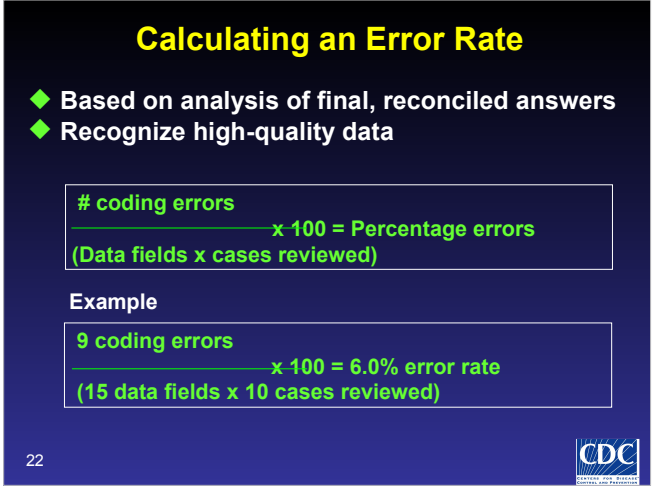


An important part of the reliability study is the reconciliation process, where the coder/abstractors work with the central registry to review the distribution of participant answers compared to the answers established by an expert panel. In contrast to reabstracting and recoding audits where the reconciliation process is between the original abstractor and the "expert" auditor, the resolution of discrepancies in a reliability study involves abstractors' discussions with the expert panel, commonly in a conference call or open meeting.

The study coordinator or reconciliation leader may indicate that data items with high concurrence rates will not be discussed during the reconciliation process. For example, if the concurrence rate on tumor behavior is 96%, there are no global issues with that data item for that case, and there is no need to spend conference time discussing it. On the other hand, if the CS Extension codes are so dispersed that no individual code has more than 30% concurrence including the preferred answer, that data item should be thoroughly discussed. Similarly, if some other code is assigned more frequently than the preferred answer, that data field should also be discussed, because something in the coding rules is leading abstractors in the wrong direction (or the expert panel has provided the wrong code).

The participants can defend their choice of codes during the reconciliation process in an attempt to persuade the expert panel to change the preferred answer, but the participants must cite specific rules or documentation as part of their discussion. In many respects, this process is as important to the education of the participants as the post-audit training sessions. In attempting to defend their choice of codes, the participant may realize why their code was incorrect and the preferred answer is correct. On the other hand, the expert panel may decide to change an answer based on evidence presented by the participants.

The entire reconciliation process is educational as the participants and the study team listen to the rationales for the answers. But right and wrong answers are not the entire reason for the reconciliation process. The discussions can include comments about the coding rules and guidelines, data collection manual documentation, and other items that can be improved as a result of the study. To cite the CS Reliability Study once more, the recommendations for clarifications in the CS Manual came largely from the study participants, who also identified areas where additional training or reference materials were needed, such as a list of immunohistochemistry tests that look for isolated tumor cells in the lymph nodes of breast cancer patients. Not only were revisions made to the CS Manual as a result of the reliability study, but educational presentations were prepared to address the issues identified in the study.



The reconciliation process may take multiple conference calls over several weeks if there are a lot of discrepancies or a lot of discussion about certain issues. At the end of the reconciliation, the answers are finalized and the participant data are run again. Sometimes the scores improve. If the finalized answers are the same as the participant's; sometimes the scores go down, if the participant's was the same as the preferred answer and it was changed during reconciliation.

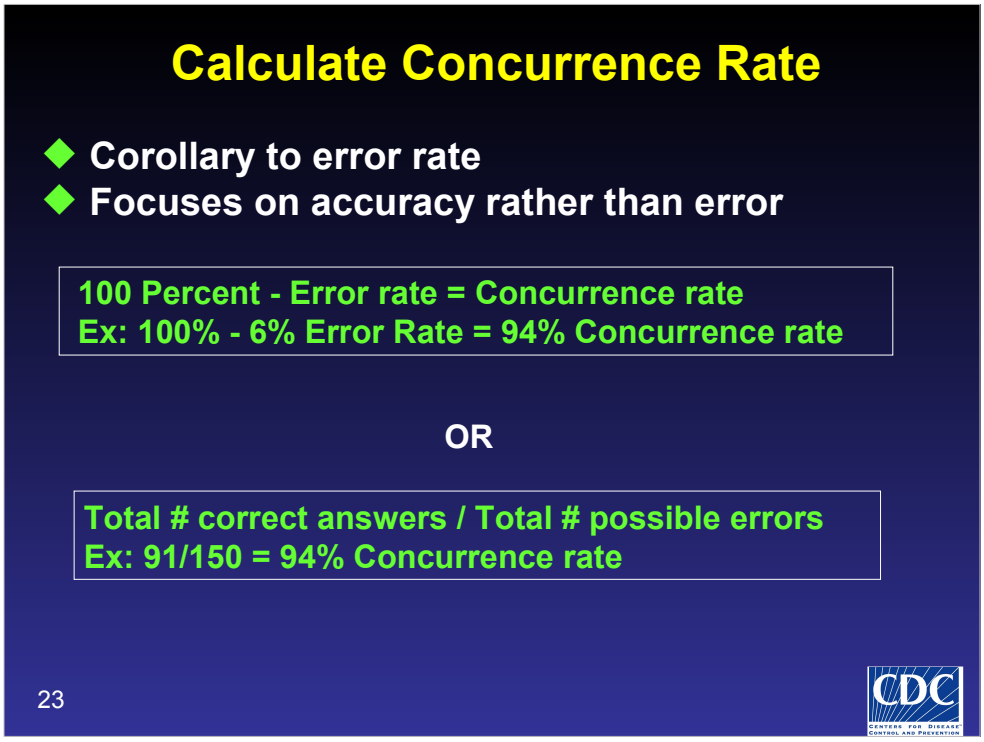
The coding differences that remain can be analyzed and reported in a variety of ways. The error rate for a reliability study is a calculated percentage similar to that of a reabstracting audit. The numerator is the number of errors. The denominator is the number of data fields reviewed per case multiplied by the total number of cases. The numerator divided by the denominator produces a decimal fraction, which, when multiplied by 100, produces an overall error rate in percent. For example, abstractor A's results identified 9 coding errors over all the cases in the study (the numerator). Fifteen data fields were abstracted for 10 cases, for a total of 150 possible errors (the denominator). The calculation would be 9 coding errors divided by 150 possible errors and multiplied by 100, for an overall error rate of 6.0%. Whether this is good, bad, or in the middle would depend on the central registry's accuracy standards as established in the study protocol. If those 9 coding errors were scattered across several data fields, the result would probably be good. If the 9 errors were concentrated in the extension field or another specific field, this would be an indication that the abstractor needed individual training in that area.

The abstractor's accuracy or concurrence rate is the corollary to the error rate, in other words, 100 (%) minus the error rate or in the example 94% percent. Alternatively, the accuracy rate can be calculated as the total number of correct answers divided by the total number of data items in the study (data fields times cases).

For a reliability study, there should be an overall error rate. In addition, the central registry may decide to subcategorize differences (after reconciliation this can be described as errors) into major, minor, and unknown-to-known. Determination of what is a major or minor difference must be part of the study protocol. In the previous example, if six of abstractor A's were minor errors, the major error rate would only be 3 of 150, or 2%. The study may also yield contextual errors; for example, if the primary site is incorrect, all of the staging and treatment codes will be incorrect. A decision would have to be made whether that will count as a single error or the total number of incorrect data fields.

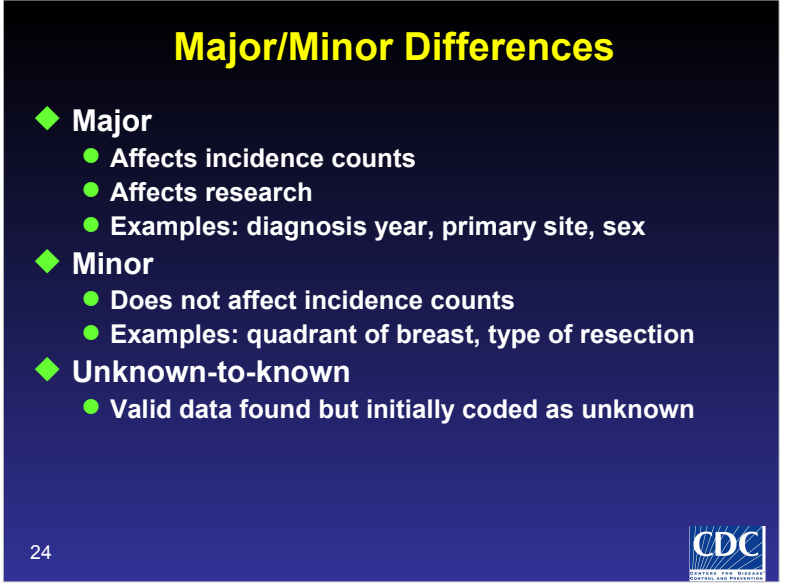
The final results can be represented statistically by comparing the results to accuracy goals for each data item, providing a measure of agreement among the data collectors. Recommendations may include further training that can be targeted to specific problems or gaps in abstractor education identified through the study or modification and clarification of rules and guidelines in data collection manuals.

Incentives based on quality assurance audits can encourage improvements in data accuracy and reliability by individual abstractors. A wide range of incentives are used by various central registries, including letters of commendation addressed to the abstractor and copied to the facility's administrators, plaques or certificates, recognition and possibly complimentary registration at local and national professional meetings, and awards of reference materials or access to products to improve the registry (such as death index access). One of the best incentives, and one which should be announced at the outset of the study, is continuing education credit for participation. This is a very practical, cost-effective way to encourage busy abstractors to take part in the study.



The abstractor's accuracy or concurrence rate is the corollary to the error rate. For example, 100 (%) minus the error rate or in the example 94%. The accuracy rate can also be calculated as the total number of correct answers divided by the total number of data items in the study (data fields times cases).

The central registry will have to decide whether to include error rates, concurrence rates, or both in the data analysis.



As noted, the outcome of a reliability study is an error rate or, conversely, a concurrence rate. However, the definition of an error is relative to how the data will be used. Some central registries categorize errors as major or minor as part of the reliability study report. For example, demographic information, particularly county/state of residence, age, sex, and race are critical measures for incidence reporting and the calculation of incidence rates. So too are primary site and tumor behavior. If these data fields can be visualized as an incidence table by primary site, any shift from one cell to another or into/out of the table itself is an important, or major error. If the incidence table involves only invasive tumors, incorrect designation of the behavior is a major error. Shifting a case from lung to prostate because the case was incorrectly abstracted as a lung primary is a major difference.

Other differences may be considered major in the context of research. if a researcher is interested in adenocarcinomas of the colon, the county/state of residence may not matter, and the age, sex, and race of the patients may be of interest for descriptive statistics. The key pieces of information in such a study would be the location of the tumor in the colon, the histologic cell type, and probably the stage and treatment information. If the research involves differences in treatment by stage, any shift of stage or type of treatment for a case would be considered a major difference. If the central registry uses major and minor differences as part of its analysis, it should clearly define the context in which they are included.

In most circumstances, a minor difference does not affect incidence counts. Even in research, the subcategory site codes of most primaries are analyzed together, such as the different lobes of the lung or the areas of the bladder. It may not matter whether a surgical procedure was an excisional biopsy or complete removal of the organ.

A difference in Collaborative Staging Extension field codes could be major or minor, depending on the derived stage that results from the codes. For example, in lung, code 10 and code 30 both map to summary stage localized, so that would probably be a minor difference. But there would be a major difference between code 10 (localized) and code 45 (regional by direct extension) in a research study based on stage at diagnosis. For breast, tumor size is an important factor for mapping to the T category in TNM, so a difference in tumor size between ranges 001–020, 020–050 and > 050 would be major differences, but differences within each range would be minor.

One other audit difference is worth noting, unknown-to-known. This occurs when the abstractor codes a data field as unknown and the expert panel finds enough information to code a specific value. Unknown-to-known differences may be a sign of inattention to detail or a signal that the abstractor needs additional training in how to interpret information in the medical record.

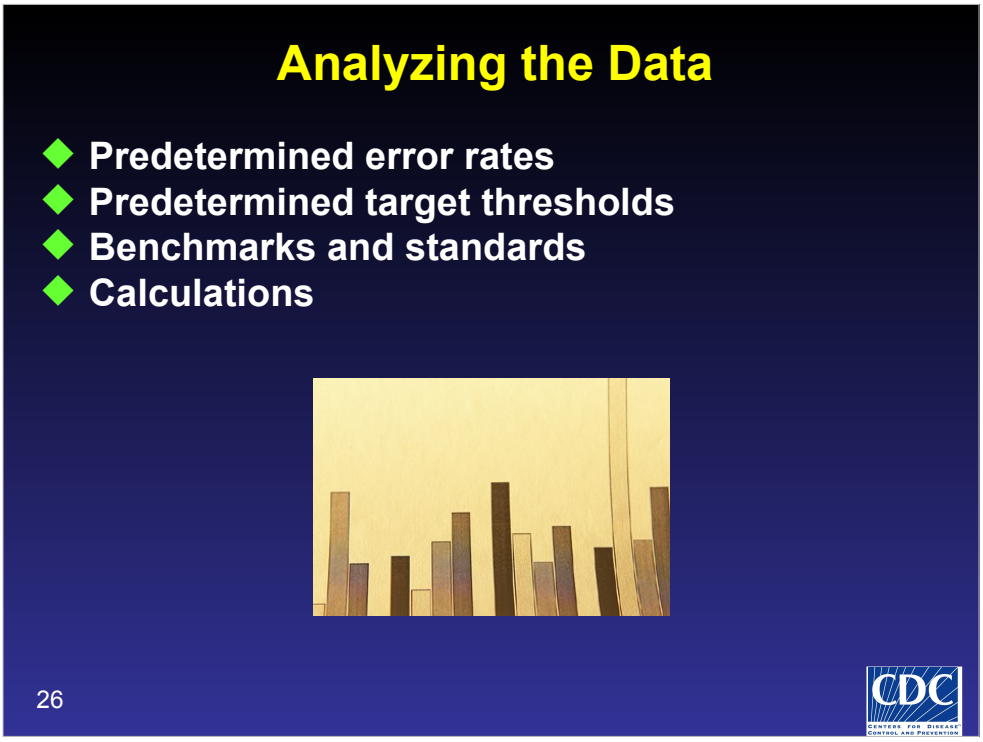
Major/Minor Difference Examples

Required Field	Major	Minor
Address at Dx - State	✓	
Race	✓	
Sex	✓	
Birth Date	Wrong ccyy	Wrong mm/dd
Date of Diagnosis (mm/dd/ccyy)	>30 days	<30 days
Primary Site	Wrong site	Wrong sub-site
Laterality	✓	
Histology Type	Wrong histology	Wrong subclass
Behavior Code	✓	
Grade		✓
Summary Stage	✓	

25

Most reliability study analysis is focused on whether any identified discrepancies affect incidence counting. In this context, most states that perform reliability studies have developed or adopted tables of major and minor differences. These examples are from an NPCR reabstracting audit. Major differences affect incidence counting and minor differences may affect other types of research. However, major/minor differences can be adjusted based on the planned use of the data.

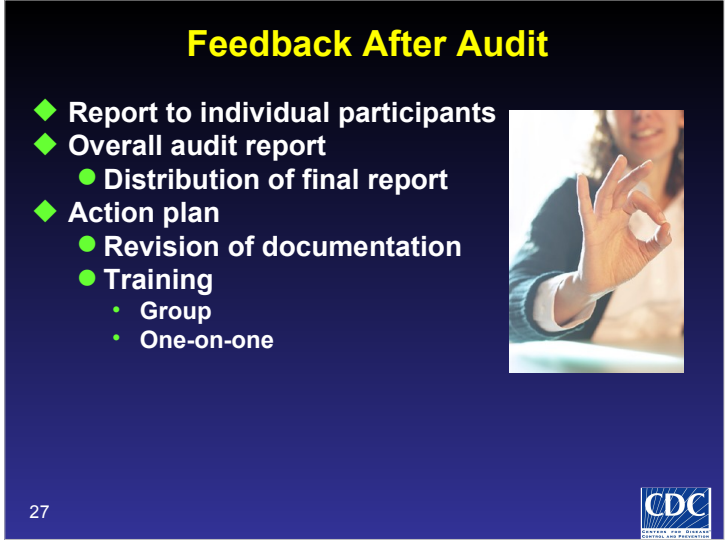
In addition, the outcomes of the study, such as incentive awards, may be based only on major errors as determined by the study design. In such cases, the minor errors are disregarded and the major error count becomes the numerator.



When the discrepancy resolution is complete, the central registry audit team can begin the analysis of findings, first by individual participant, and then in aggregate. The design of the study will affect how the data are analyzed. For example, the audit team may have made a decision to compare all participants against a standard, such as a threshold for accuracy. This might be considered a "pass-fail" type of standard—if the participant exceeds the threshold, he or she passes, regardless of how "perfect" the answers are. If the threshold is not met, the participant fails, regardless of how many errors were made. Another option might be considered a grading system with ranges for error rates on a scale of one to five stars, A-B-C-D-F, or poor-satisfactory-excellent. With any of these options, the ranges would be pre-determined as part of the study design. Either way, the central registry must decide what the acceptable quality level should be in advance of the study itself.

Benchmarks will vary according to the type of audit and its purpose. For a reliability study, the benchmark is usually established by the central registry conducting the study, since there are no published federal standards for data accuracy. For data quality, benchmarks can also be obtained from published reports developed by other central registries and the Commission on Cancer's National Cancer Data Base, as well as by reviewing journal articles and documents issued by NAACCR.

The calculations may be considerably more detailed than just a simple percent of data items determined to be "wrong." More in-depth calculations may be reported, such as the percent of errors by data "cluster" (demographics, tumor identification, staging, treatment), by type of error (major, minor, unknown-to-known), or by primary site. If available, data from previous reliability studies of similar nature can be included in the current report to identify improvement or further deficiencies.



The last step in a reliability study is feeding the results back to the participants in the study and creating an aggregate report. This is more than just common courtesy for those who were involved in the audit, this is closing the loop on data quality assurance. Unless the participants are notified of the findings, they will not be able to take the appropriate actions to correct the errors and develop strategies for education or process improvement that hopefully result in a lower error rate during the next study. And eventually, there will be another study to monitor whether the corrections were effective.

Final results can be provided in a technical document. Because of the amount of time each participant has literally donated to the reliability study, a fairly detailed analysis of the results would be advisable. Such an analysis could include a description of the study and when it was conducted, discussion of the testing points for each case, graphic displays of distribution of answers, summary of comments and recommendations based on the reconciliation discussions, and perhaps even specific educational points relevant to the data item. It would be possible to create an overall, detailed, summary report and provide a separate document containing the individual participant's results, including the number of data items reviewed per case, number of correctly abstracted data items, and number of incorrectly abstracted data items, together with an overall accuracy rate and the formula for calculating that rate.

It would be helpful to summarize the errors by major and minor categories and discuss any pattern of errors. A cover letter that contains thanks for cooperation and congratulations for a good job (if applicable) should be sent to the abstractor with a copy or separate letter to the hospital administrator. The overall audit report will be combined data from all the participants. In this way, patterns of error rates can be identified and plans made for further group trainings.

Another question to be addressed is the distribution of the final report. Should it be published on the central registry's Web site? Who, besides the quality assurance team should see it? Certainly researchers using the data should be informed of any findings of concern, such as weaknesses in the accuracy and reliability of certain data fields or the consistency of staging and treatment codes. Keep in mind, however, that a report on reliability does not necessarily address the accuracy of the entire database, but rather assesses the consistency with which codes were assigned to the cases in the study.

Finally, the overall report needs an action plan. The action plan is based on interpretation of the analysis and the overall experiences of the study team and the participants. For a reliability study, did the final analysis indicate that only certain fields (or groups of fields like the staging section) were problem areas? If so, the action plan should describe training efforts, either for all data collectors in the state or for individual data collectors, as well as inservice training as needed for central registry staff. Did the audit reveal that there were areas in the central registry's documentation that were not clear? If so, the action plan should describe the areas of the coding manual or the reportable list and reporting requirements that should be clarified. Did the audit indicate that the protocol was not clear enough to the participants? If so, revise the protocol now while it is still fresh in mind, so that the improvements won't be forgotten the next time a similar audit is done in the future. If the state central registry participated in a national study conducted by a standards setter, the overall results of the state's participants should be shared with the central registry, which can then analyze the results and identify any issues to be addressed through state training programs.

This reliability study is complete. Now it's time to step back, take a deep breath, and begin the quality improvement cycle once again.

Central Registry Response

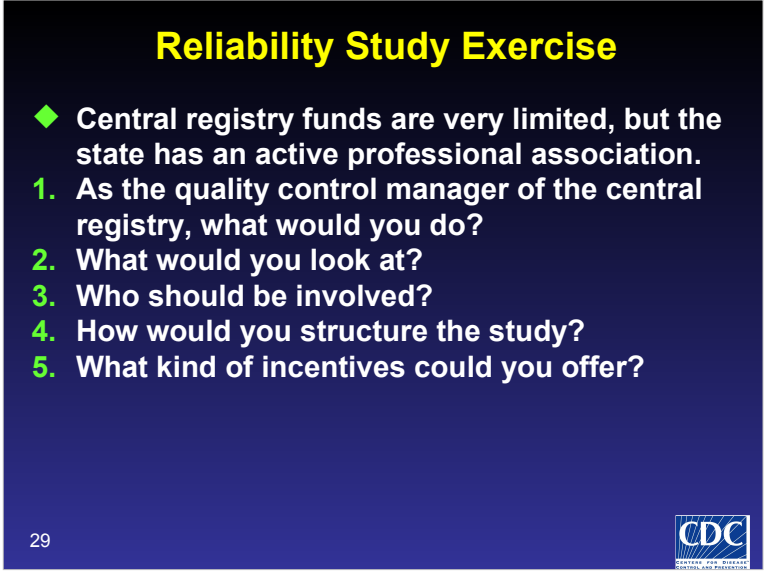
- Follow up on action plan
  - Identify education needs
  - Revision of documentation
- Training
  - Develop and conduct training
- Participant recognition
  - Letter of commendation
  - Conference scholarship

28

The central registry should follow up on cited recommendations and actions.

This may include the development of further training that can be targeted to specific problems or gaps in abstractor education identified through the study or modification and clarification of rules and guidelines in data collection manuals.

Incentives based on results can encourage improvements in data accuracy and reliability by individual abstractors. A wide range of incentives are used by various central registries, including letters of commendation addressed to the abstractor and copied to the facility's administrators, plaques or certificates, recognition and possibly complimentary registration at local and national professional meetings, and free reference materials or access to products to improve the registry.



[Read the slide to establish the discussion points.]

Possible answers:

1A. Do some comparisons on the incoming data to determine whether there is a pattern of specific abstractors or specific facilities reporting the new fields as blank or unknown.

1B. Contact the facilities with the highest rates of blanks or unknowns and question if the information is available.

1C. Decide to perform a targeted reliability study on the new data fields.

1D. Decide to perform a targeted reliability study on the fields for which the rules shifted from COC to SEER standards.

1E. Combine the targeted reliability studies to include all the changes that took place when the new data collection manual was published.

2A. Look at the new data fields.

2B. Look at any data field affected by the rules shifts. These might be sequence of primary, case reporting (VIN, VAIN, AIN), date fields, or any others identified as changes in the previous year's trainings.

3A. In particular, data collectors who did not attend the previous year's trainings (if this can be determined), and abstractors at the facilities with missing or unknown codes in the new data fields.

3B. This would be an ideal time to assess all abstractors in the state regarding the changes in the new manual.

4A. There are a variety of ways to structure a study. Full cases could be posted on an Internet site as a full-blown reliability study, or brief case scenarios asking very specific data questions. This could be presented at a central registry or registrar association workshop. One or two cases at a time could be included in a central registry or state association newsletter. The point is to bring attention (again) to the changes in the manual.

5A. Continuing education credits for participating in a study or submitting answers to case scenarios printed in a newsletter.

5B. For a full reliability study, award certificates (gold level accuracy, silver and bronze level accuracy) or some other recognition.

Reliability Study Exercise

- At the beginning of last year, the central registry published a new data collection manual that changed some of the abstracting rules from COC standards to SEER standards. Training sessions were held around the state, but not everyone attended. In addition, several new cancer risk data fields became required.
- The new requirements have been in place for a full calendar year and registrars in the state have been abstracting under the new rules for about six months. Data are now coming in, and you've noticed that many of the new fields are filled with unknowns.

30

Let us try to focus what we've discussed on an issue raised by researchers using central registry data.

[read the slide to explain the scenario]

Resources

1. Dryden M and Brogan K. Quality Control. Chapter 20 in Menck H et al., *Central Cancer Registries: Design, Management and Use, second edition*. Kendall Hunt Publishing Co., 2007.

2. Unpublished materials provided by National Program of Cancer Registries

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

For information about CDC's
Cancer Prevention and Control Programs
and the
National Program of Cancer Registries

Please visit www.cdc.gov/cancer/npcr